Relating SARS-CoV-2 shedding rate in wastewater to daily positive tests data: A consistent model based approach

M. Petala[1], M. Kostoglou[2], Th. Karapantsios[2], Ch. Dovas[3], Th. Lytras[4,5], D. Paraskevis[6], E. Roilides[7], A. Koutsolioutsou-Benaki[8], , G. Panagiotakopoulos[4], V. Sypsa[6], S. Metallidis[9], A. Papa[10], E. Stylianidis[11], A. Papadopoulos[12], S. Tsiodras[6], N. Papaioannou[3]

[1]Laboratory of Environmental Engineering & Planning, Department. of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, 54 124, Greece

[2]Laboratory of Chemical and Environmental Technology, Dept. of Chemistry, Aristotle University of Thessaloniki, 54 124 Thessaloniki, 54124, Greece

[3]Faculty of Veterinary Medicine, School of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

[4]National Public Health Organization, Athens, Greece

[5]European University Cyprus, Nicosia, Cyprus

[6]National and Kapodistrian University of Athens, Athens, Greece

[7]Infectious Diseases Unit and 3rd Department of Pediatrics, Aristotle University School of Health Sciences, Hippokration Hospital, Thessaloniki, 54642, Greece

[8]Department of Environmental Health, Directory of Epidemiology and Prevention of Non Communicable Diseases and Injuries, National Public Health Organization, Athens, Greece

[9]Department of Haematology, First Department of Internal Medicine, Faculty of Medicine, AHEPA General Hospital, Aristotle University of Thessaloniki, Thessaloniki, 54636, Greece

[10]Department of Microbiology, Medical School, Aristotle University of Thessaloniki, Thessaloniki, 54124 Greece

[11]School of Spatial Planning and Development, Faculty of Engineering, Aristotle University of Thessaloniki, 54124, Greece

[12]EYATH S.A., Thessaloniki Water Supply and Sewerage Company S.A., Thessaloniki, 54636, Greece

ABSTRACT

During the COVID-19 pandemic, wastewater-based epidemiology (WBE) has been engaged to complement medical surveillance and in some cases to also act as an early diagnosis indicator of viral spreading in the community. Most efforts worldwide by the scientific community and commercial companies focus on the formulation of protocols for SARS CoV-2 analysis in wastewater and approaches addressing the quantitative relationship between WBE and medical surveillance are lacking. In the present study, a mathematical model is developed which uses as input the number of daily positive medical tests together with the highly non-linear shedding rate curve of individuals to estimate the evolution of virus shedding rate in wastewater along calendar days. A comprehensive parametric study by the model using as input actual medical surveillance and WBE data for the city of Thessaloniki (~700,000 inhabitants, North Greece) during the outbreak of November 2020 reveals the conditions under which WBE can be used as an early warning tool for predicting pandemic outbreaks. It is shown that early warning capacity is different along the days of an outbreak and depends strongly on the number of days apart between the day of maximum shedding rate of infected individuals in their disease cycle and the day of their medical testing. The present data indicate for Thessaloniki an average early warning capacity of around 2 days. Moreover, the data imply that there exists a proportion between unreported cases (asymptomatic persons with mild symptoms that do not seek medical advice) and reported cases. The proportion increases with the number of reported cases. The early detection capacity of WBE improves substantially in the presence of an increasing number of unreported cases. For Thessaloniki at the peak of the pandemic in mid-November 2020, the number of unreported cases reached a maximum around 4 times the number of reported cases.

HIGHLIGHTS

- Model estimates viral load evolution in wastewater from infected people dynamics
- Identifying actual conditions for which WBE can be used as an early warning tool
- Early warning capacity increases with an increasing number of unreported cases
- In Thessaloniki Nov20 outbreak, the early warning capacity of WBE was about 2 days
- In Thessaloniki Nov20 outbreak, unreported cases were up to 4 times reported cases

INTRODUCTION

In the COVID-19 pandemic, wastewater-based epidemiology (WBE) has become an important tool, supplementing public health surveillance, in the hands of scientists, medical experts, and state officials to evaluate the spread of SARS-CoV-2 in the community [(la Rosa et al., 2020; Medema et al., 2020b; Wurtzer1 et al., n.d.)]. This was more so after the Centers for Disease Control and Prevention (CDC) adopted wastewater disease surveillance as a valid monitoring tool and provided guidance and recommendations for the selection and application of testing methods (CDC, 2021). The latter cover issues from sample collection and sample processing to RNA measurement and use of laboratory controls for the estimation of the performance of the applied methods and of data quality. Laboratory controls deal with difficulties arising from the chemical and biological variability of wastewater quality across different places and over time, e.g. due to season, weather, human activities etc., and also cope with problems of proper RNA extraction and quantification, elimination of inhibition and contamination of reagents.

The so-called matrix recovery controls suggested by CDC refer exclusively to the amount of virus lost during sample processing. However, it is well known that viruses strongly adsorb, and so get inaccessible ("lost"), in the pores of solid particles suspended in wastewater in sewerage networks (Sellaoui et al., 2020; Ye, 2018). Therefore, if WBE studies aim to quantify the virus shedding rate at the source (households) from analysis of samples taken at the entrance of wastewater treatment plants

75 (WTP) then it is of paramount importance to appraise the amount of virus lost in sewerage networks.
76 Determining the extent of recovery, as suggested by CDC, by spiking techniques in samples obtained at
77 the entrance of WTP is incapable of describing the amount of virus lost in sewerage networks. This is
78 because such samples have traveled in sewerage pipes for hours and so have (most of) their adsorption
79 sites saturated with adsorbed species. Desorption of spiked material would have been completely
80 different if a sample had been taken fresh at a shedding spot, i.e., from the sanitary plumbing system of
81 a building.

82 Recently, a mathematical model has been proposed to account for virus loss in sewerage
83 networks by means of adsorption of virus on porous particles suspended in wastewater(Petala et al.,
84 2021). To do so, the model rationalizes the SARS-CoV-2 concentration in samples taken at the entrance
85 of WTP with respect to certain quality characteristics of wastewater samples. Rationalization is based on
86 rigorous physicochemical phenomena in adsorption (and not on a statistical dependence of virus
87 concentration on wastewater parameters), also including the effect of large-scale topological complexity
88 of actual sewage networks. At the examined period of time (April to June 2020), the rationalized
89 decreasing shedding rate was in agreement with the observed clinical conditions, contrary to the non-
90 rationalized data which showed a different picture. Yet, even with rationalized data it is still difficult to
91 make the critical step ahead and associate the virus shedding rate with the number of cases.

92 There are several reasons for the discrepancy between viral shedding rate in wastewater and
93 number of infected people (cases) reported by public health surveillance. Limitations in the capacity of
94 medical testing, strict criteria for the application of medical testing and parts of the population being
95 reluctant to seek medical care are responsible for underdiagnosis of cases, mainly asymptomatic cases
96 or patients with mild symptoms who are not tested. These factors lead to reporting a lower number of
97 cases than the actual one. On the other hand, poor recovery in the collected wastewater samples and
98 virus loss in sewage networks, along with ineffectiveness to properly rationalize for the latter, lead to a

99 smaller measured concentration of virus in the collected samples and, thus, to a lower estimated viral
100 shedding rate than the actual one.

101 The time delay between the onset of patients' viral shedding in infected people and symptoms
102 onset that prompts patients being tested, is another reason for discrepancy. Several reports in literature
103 indicate a time delay up to 8 days between wastewater signal and infected cases reported in public
104 health surveillance system (e.g.(D'Aoust et al., 2021; Nemudryi et al., 2020; Peccia et al., 2020)). This is
105 in line with the reported incubation period of around 6 days from exposure/infection to onset of
106 symptoms of COVID-19 (Guan et al., 2020; Lauer et al., 2020; Li et al., 2020). Therefore, the discrepancy
107 between the onset of viral shedding and the onset of symptoms is actually an advantage that might
108 explain the early detection capacity of WBE.

109 Another important reason of discrepancy between daily reports of wastewater viral titers and
110 daily reports of new confirmed cases is the cumulative character of viral shedding rate in wastewater
111 during the disease. Several clinical studies 14–16 showed that viral shedding is not uniform during the
112 disease, but there is a maximum shedding rate in the very acute phase of infection, probably even
113 before respiratory symptoms appear, being followed by an exponential decline in subsequent days. If
114 one considers that shedding lasts at least 3-4 weeks after the inception of symptoms 14,15. it is
115 apparent that daily wastewater viral titers correspond more to the cumulative shedding of infected
116 people rather than to the shedding of the daily new cases. It must be stressed here that the evidence
117 from clinical studies is limited, and only for samples collected from hospitalized patients, thus after
118 symptoms onset, so one should be extremely careful in estimating shedding rates at the first days right
119 after infection.

120 If one aims to associate wastewater surveillance data with public health surveillance data, the
121 different sources of bias between these data should be also taken into account. Laboratory practice
122 shows that variation in wastewater data is considerably high from one day to the next because of the

123 many experimental parameters that are involved, e.g., individual's shedding rate (orders of magnitude
124 variation among infected people(Zheng et al., 2020); sampling protocol, sample concentration
125 techniques, RNA extraction methodology, normalization with respect to population indicators and flow
126 rate, inhibition assessment in RNA recovery, RNA quantification, etc. On the other hand, public health
127 surveillance is susceptible to a systematic bias because of limitations in the capacity of medical testing
128 along with the unpredictable human behavior under stressful conditions. In principle, one should be
129 concerned about the correctness of both wastewater viral measurements and clinical testing data.

130 The goal of this work is to setup a methodology for the consistent comparison between the
131 number of infected people reported by clinical surveillance and SARS-CoV-2 RNA concentration in
132 wastewater samples. To do so, a model is developed for the estimation of the evolution of virus
133 shedding rate in a sewage system based on the daily positive medical tests. Apart from corrections
134 based on the laboratory controls suggested by CDC, virus shedding rate data are further rationalized
135 with respect to physicochemical parameters of wastewater to account for virus loss by adsorption to
136 sewage solids (Petala et al., 2021). To increase accuracy, clinical testing data are based on the date of
137 specimen collection and not on the reporting date. The model is first developed in a generalized
138 continuous time regime and then it is transformed to its discrete counterpart dictated by the daily basis
139 data.

140

141 **METHODS**

142 **Sampling**

143 Wastewater samples were collected at the exit of Thessaloniki's (a city at North Greece) main sewerage
144 pipe, right before the entrance at the wastewater treatment plant of the city, as described
145 elsewhere(Petala et al., 2021). This plant serves an estimated population of about 700,000 people. The

146 present work reports 24-hours composite samples (1L each) taken three times per week (Monday-
147 Wednesday-Friday) from October 5th, 2020 until January 6th, 2021. This period includes the second wave
148 burst of COVID-19 for Thessaloniki in November 2020 which was the worst ever in the whole country.
149 Typical range of values of the physicochemical parameters of wastewater samples for the examined
150 period are displayed in Table 1 (supplementary materials). These parameters were employed in the
151 rationalization of the measured viral concentration with respect to quality characteristics of wastewater
152 according to Petala et al. (Petala et al., 2021).

153 **Virus concentration**

154 Upon wastewater collection, 200 mL of each wastewater sample was centrifuged at 4000xg for 30 min
155 to remove particles and pH of the supernatant was adjusted to 4 using a solution of 2.0 M HCl. An
156 aliquot of 40 mL supernatant was then passed through an electronegatively charged surface of 0.45 µm-
157 Ø47 mm cellulose nitrate HA membrane ((HAWP04700; Merck Millipore Ltd., Tullagreen, Ireland).
158 Filtration was performed using a magnetic funnel mounted on a glass filtration flask (Pall Corporation)
159 Filtration step was conducted in triplicate and membranes were stored in 15 mL falcon tubes for further
160 processing of RNA extraction and virus quantification.

161 **RNA extraction and virus quantification**

162 For each sewage sample, three electronegative membranes were individually subjected to phenol-
163 chloroform-based RNA extraction process (Chaintoutis et al., 2019) coupled with magnetic bead binding.
164 Each membrane filter was rolled into a Falcon™ 15 mL conical centrifuge tube with the top side facing
165 inward. The following were added sequentially, a) 900 µL of guanidinium isothiocyanate-based "Lysis
166 buffer I" [5M guanidinium isothiocyanate, 25mM EDTA, 25mM sodium citrate (pH 7.0), 25mM
167 phosphate buffer (pH 6.6)] containing 1% N-Lauroylsarcosine, 2% Triton X-100, 2% CTAB and 2% PVP, b)
168 18µl β-mercaptoethanol, c) 300µl H2O, mixed thoroughly by inversion, and each tube was incubated at

169 4 °C on horizontal rotator (50 rpm) for 10-30 min. 1200 µl of "Lysis buffer II" were added [prepared by
170 mixing 152.5gr guanidinium hydrochloride, 31.25 ml of 2M acetate buffer (pH 3.8) and water-saturated
171 phenol stabilized (pH4), up to 500ml final volume], followed by incubation 10 min / RT on horizontal
172 rotator (150 rpm). The liquid phase was transferred on a 2-mL microcentrifuge tube and clarified by
173 centrifugation (21,000×g, 5 min, 4 °C). 1600 µl were transferred to a new tube, 200 µl chloroform-
174 isoamyl alcohol (24:1) were added and shaken vigorously for 30 s followed by incubation (–20 °C, 30
175 min) and centrifugation (21,000×g, 10 min, 4 °C). 800 µl of the upper aqueous phase were transferred
176 and mixed with 667µl isopropanol and 20µl of magnetic beads (IDEXX Water DNA/RNA Magnetic Bead
177 Kit), followed by incubation 15 min / RT on horizontal rotator (150 rpm). The beads were washed
178 according to the manufacturer's protocol, using the magnetic extraction reagents. Elution of RNA was
179 done in 100 µl buffer followed by filtering using OneStep PCR inhibitor removal kit (Zymoresearch).

180 SARS-CoV-2 quantified RNA (1x 106.3 genomic copies) from human clinical samples was added to a
181 subset of concentrates to estimate the recovery efficiency and reproducibility of the RNA extraction
182 procedure. Similarly, heat inactivated SARS-CoV-2 from human clinical samples was spiked (1x 107 viral
183 particles) in a subset of sewage samples to assess the recovery and reproducibility of the virus
184 concentration procedure.

185 Each RNA extract was subjected to real-time RT-PCR for SARS-CoV-2 quantification in triplicates. Two
186 primer/probe sets were utilized: the N2 set from CDC that targets the nucleocapsid (N) gene and the set
187 targeting the genomic region that encodes the E protein (Corman et al., 2020). The assays were
188 performed on a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA,
189 USA). Reactions were considered positive if the cycle threshold was below 40 cycles. Calibration curves
190 were generated using the synthetic single-stranded RNA standard "EURM-019" (Joint Research Centre,
191 European Commission). The possible presence of RT-PCR inhibitors in each RNA extract was assessed in

192 duplicates using spiked EURM-019 included at ~1000 copies in additional RT-PCR reactions. Inhibition
193 was expressed as % reduction in reported copy number, compared to the sum of spiked EURM-019
194 copies and the mean value of measured SARS-CoV-2 genomic copies in the non-spiked RNA. SARS-CoV-2
195 viral load in each sewage sample was expressed as mean ± standard deviation genome copies per liter,
196 after correcting for RT-PCR inhibition if present, and recovery efficiencies (virus concentration and RNA
197 extraction). Precision of each individual sewage sample quantification was assessed using the coefficient
198 of variation (CV) of the estimated SARS-CoV-2 viral load of the three electronegative membranes
199 processed. We set our precision threshold at 35% CV for each sewage sample measurement.

200 **Epidemiological data**

201 Daily numbers of COVID-19 infected people reported in the city of Thessaloniki, adjusted to specimen
202 collection date, were obtained from the National Public Health Organization (National Public Health
203 Organization, 2020). These data reflect residents found positive when tested in public and private
204 laboratories. Data are presented for the period between September 1st, 2020 and January 6th, 2021
205 (original data are provided in supplementary materials).

206 **Problem formulation**

207 A model is developed for estimating the evolution of virus shedding rate to sewage system based on the
208 official (announced by the state) results of daily positive tests (cases), registered by the date of
209 specimen collection and not by the date of public reporting. There is typically a delay of few days
210 between specimen collection and public reporting. The model is first developed in a generalized
211 continuous time regime and then it is transformed to its discrete counterpart dictated by the daily basis
212 data.

Let us denote as f(t) (t: calendar day) the evolution of positive test counts density with f(t)dt being the number of positive tests in the time period between t and t+dt. The first step is to transform the function f(t) to the function F(t) which denotes the total number of reported infected people at time t. It is essential to stress here that F(t) includes people that at time t either had already a positive test or they are in the first days after infection, so do not have symptoms, but still they shed virus and will be tested positive later after their onset of symptoms. The effect of infected but unreported people, such as asymptomatic ones and those not tested because of mild symptoms, is dealt with later.

In order to proceed let us consider the course of the disease of an infected person: Infection starts (disease incidence) at day $\tau=1$, detection occurs (specimen collection) at day $\tau=\tau_d$ and end of viral shedding occurs at day τ_e. The number of days for detection and end of shedding are in general not the same among cases but they show a dispersion which can be described by the respective probability density functions $P_d(\tau_d)$ and $P_e(\tau_e)$. It is understood that both functions take non-zero values only in restricted domains of their arguments. The limits of these domains are denoted as τ_{d1}, τ_{d2} and τ_{e1}, τ_{e2}, respectively. The functions P_d and P_e satisfy the conditions:

$$\int_{\tau_{d1}}^{\tau_{d2}} P_d(\tau_d)d\tau_d = 1 \qquad (1a)$$

$$\int_{\tau_{e1}}^{\tau_{e2}} P_e(\tau_e)d\tau_e = 1 \qquad (1b)$$

$$\int_{\tau_{d1}}^{\tau_{d2}} \tau_d P_d(\tau_d)d\tau_d = \tau_{d,av} \qquad (1c)$$

$$\int_{\tau_{e1}}^{\tau_{e2}} \tau_e P_e(\tau_e)d\tau_e = \tau_{e,av} \qquad (1d)$$

where the subscript "av" denotes the average value of the corresponding variable. Taking into account the above definitions, it can be shown that the functions F(t) and f(t) can be related to each other as:

$$F(t) = \int_{\tau_{d1}}^{\tau_{d2}} \int_{\tau_{e1}}^{\tau_{e2}} P_d(\tau_d) P_e(\tau_e) \int_{t+\tau_d-\tau_e}^{t+\tau_d} f(x) dx d\tau_e d\tau_d \qquad (2)$$

This relation is actually the mathematical expression of the statement that infected persons at time t include those already detected and those to be detected in the next days. In principle, a multiplication of this number by the average shedding rate per person would give the required global shedding rate evolution function R(t). Such averaging over all infected individuals would have been sufficient for data reduction purposes only if the shedding rate of individuals during their disease cycle had a constant value. However, as already explained, this is not true since there is a strong variation in virus concentration in stool not only among infected individuals but also across the days of their disease cycle.

Let us denote this variable function of the daily shedding rate per person as S(τ), (τ: day of the disease onset). It has been shown that not only S is not a constant, but on the contrary, it is a strong function of its argument τ. In order to incorporate the effect of function S(τ) to the global shedding rate R(t), knowledge of the distribution of disease days among the infected population is required. The corresponding function is denoted as F*(τ,t) and represents the density function of the distribution of disease days, τ, at time t. It is fruitful to decompose the function F*(τ,t) into the product of the total number of infected people and the probability density function of the days of infection as F*(τ,t)=F(t)g(t,τ) where g(t,τ) is a probability density function with respect to τ and satisfies the condition:

$$\int_0^{\tau_{e2}} g(\tau, t) d\tau = 1 \qquad (3)$$

It can be shown that the function g can be derived as

$$g(\tau,t)=\frac{1}{F(t)}\int_{\tau_{d1}}^{\tau_{d2}}\int_{\tau_{e1}}^{\tau_{e2}}P_d(\tau_d)P_e(\tau_e)f(t+\tau_d-\tau)U(\tau_e-\tau)d\tau_e d\tau_d \qquad (4)$$

where U is a step function taking the values 0 for negative and 1 for positive argument. The domain of definition of the above expression is from $t=\tau_{e2}-\tau_{d1}$ to $t=T-\tau_{d2}$ in the case of f(t) known from t=1 to t=T.

Parametrization of the function S(τ)

In order to proceed, one should know the function of virus shedding rate in stool per person and per day of the disease. A first question is how the onset of infection (i.e., time τ=1) is defined. There are several possibilities for this definition but in the present context the most relevant one is that τ=1 day is the first day of non-zero shedding of virus in the stool of a person.

There are only a few available clinical studies in literature (e.g. Wölfel et al.(Wölfel et al., 2020), Huang et al.(Huang et al., 2020); Tan et al.(Tan et al., 2020)) referring to the kinetics of SARS-CoV-2 shedding rate in stool during the course of disease (from τ=1 to τ_e). To our best knowledge, from those studies only the work of Wölfel et al.(Wölfel et al., 2020) reports actual viral concentrations whereas all other studies refer to Ct values from molecular analysis. It must be mentioned that these studies present data exclusively from hospitalized patients, thus from people presenting moderate to severe symptoms. Apparently, it is not easy to collect stool samples from infected persons prior the onset of symptoms, and as a result, asymptomatic or mild cases are not registered in clinical studies. A first common observation in these studies is that virus concentration in stool among infected persons vary by several orders of magnitude. Therefore, here an average shedding function per person will be employed and the significance of variations among individuals will be discussed on statistical terms. Interestingly, many past studies assumed in addition a uniform with respect to time shedding rate despite the orders of

magnitude variability across the course of the disease (Ahmed et al., 2020; Gonzalez et al., 2020; Medema et al., 2020a; Saththasivam et al., 2021). This assumption simplifies the algebra of the problem enormously but it is incorrect.

A thorough fitting procedure of the virus titers data in stool of Wölfel et al.(Wölfel et al., 2020), led to a peculiar two-steps model(Miura et al., 2021): a first step with zero viral shedding, where virus load simply accumulates in infected hosts up to a maximum concentration reached at the day of symptoms onset, and a subsequent second step characterized by viral shedding at an exponentially decreasing concentration over the days of the disease. This fitting yields a Gamma function which degenerates to an exponential one. Evidently, it is quite arbitrary to assume that there is an initial viral accumulation period up to a maximum concentration without any shedding at all.

Here, an even more general parameterization is introduced. The average over the infected persons function is represented as a product of the following factors: (1) the average stool amount produced by a person per day, A (g_{stool}/day), (2) the maximum in time (averaged over infected individuals) virus concentration in stool, B (g_{virus}/g_{stool}) and (3) a time distribution function $s(\tau)$ of this concentration. This distribution is assumed to consist of an exponential increase from a minimum initial value up to a maximum value being followed by an exponential decrease down to a minimum final value. These minimum initial and final values of the distribution are assumed to be 1% of the maximum value. As a result, the whole distribution spans shedding rates over two orders of magnitude. This assumption may be changed to any other option, e.g., to 1‰ (three orders of magnitude span), but the approach remains the same. However, a two orders of magnitude span represents adequately most of the shedding of infected individuals during a typical course of the disease, so it is adopted herein. A new parameter is introduced, τ_a, which denotes the value of τ days at which the maximum virus

concentration in stool, i.e. maximum shedding rate, occurs. Summarizing, the above the function $S(\tau;\tau_a,\tau_e)$ is given as:

$$S(\tau;\tau_a, t_e)=ABs(\tau;\tau_a,\tau_e) \qquad (5)$$

where $s(\tau;\tau_a,\tau_e)$ is

$$s=0.01\exp(4.6\tau/\tau_a) \qquad \tau<\tau_a \qquad (6\alpha)$$

$$s=\exp(4.6(\tau-\tau_a)/(\tau_e-\tau_a)) \qquad \tau>\tau_a \qquad (6b)$$

The shape of the function $s(\tau)$ for several pairs of (τ_a,τ_e) - (6,32), (12,32), (9,25) - is shown in **Figure 1**. The particular parametrization of the function S proposed in the present work is more realistic than the Miura et al. (Miura et al., 2021) model, as it incorporates viral shedding even in the early days of the disease and before the peak maximum (peak) viral concentration in stool is reached. Such early shedding in stool is in line with clinical studies reporting patients with shedding in oropharyngeal swabs (He et al., 2020) and gastrointestinal symptoms (Siegel et al., 2020) a few days before the appearance of respiratory symptoms. Nevertheless, one should keep in mind that respiratory shedding is only a proxy for fecal shedding. In addition, the proposed function S is very attractive because there is one to one correspondence between the parameters and major features of the function. The height (amplitude) of S is determined by the product AB, its time length is determined by the parameter τ_e and its skewness by the parameter τ_a. In particular, the closer τ_a is to $\tau_e/2$ the smaller the skewness is. It must be stressed that the proposed parametrization of S does not require explicit information on the duration of the incubation period after infection nor on the day of symptoms onset. What actually matters in the context of the present analysis is the day of the maximum shedding rate as this is described by the parameter τ_a. Some clinical studies indicated that this maximum might occurs at the end of the incubation period which is taken also as the day of symptom onset (Huang et al., 2020; Miura et al.,

2021; Tan et al., 2020). However, in the present formulation this day is flexible and can be anytime from the moment of infection to the end of the shedding period (i.e., the active days of the disease).

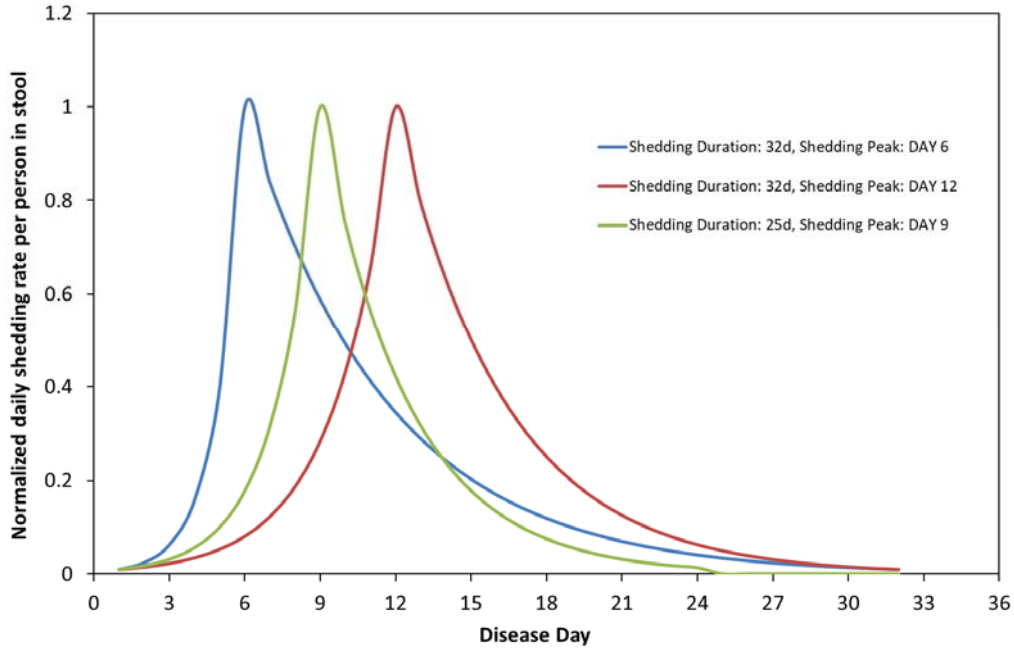


Figure 1. Time distribution of normalized shedding rate per person in stool at different shedding durations and shedding peak days.

The final expression for the global shedding rate is taken by integrating the individual's shedding rate over the days of the disease.

$$R(t) = \int_{0}^{\tau_{e2}} \int_{\tau_{d1}}^{\tau_{d2}} \int_{\tau_{e1}}^{\tau_{e2}} P_d(\tau_d) P_e(\tau_e) f(t + \tau_d - \tau) S(\tau; \tau_a, \tau_e) U(\tau_e - \tau) d\tau_e d\tau_d d\tau \qquad (7)$$

The development up to now is rather complex and includes several unknown probability density functions. In the absence of any information about them, it is convenient to consider them as Dirac delta functions. In this way, we assign to each distribution its average value. This approach not only simplifies considerably the mathematical problem but it also allows -through sensitivity analysis- to assess bounds

on the effect of using a different distribution than the Dirac delta function. This is based on the principle that any secondary feature of a distribution has much smaller effect on the result than its average value. By considering the relations $P_d(\tau_d)=\delta(\tau_d-\tau_{d,av})$ and $P_e(\tau_e)=\delta(\tau_e-\tau_{e,av})$ (where δ denotes the Dirac delta function) and substituting them in equations (2), (4), (7) leads after some algebra to (the subscript "av" is dropped in the following for clarity):

$$F(t) = \int_{t+\tau_d-\tau_e}^{t+\tau_d} f(x)dx d\tau_e d\tau_d \qquad (8)$$

$$g(\tau, t) = \frac{f(t+\tau_d-\tau)U(\tau_e-\tau)}{F(t)} \qquad (9)$$

$$R(t) = F(t)\int_0^{\tau_e} g(\tau, t)S(\tau;\tau_a,\tau_e)d\tau \qquad (10)$$

The above set of equations is discretized to be compatible with the present data. These data refer to daily values of virus shedding rate so a finite volume discretization is followed here. Let us denote as f_i the number of positive medical tests at day i (measurement period i=1 to N), F_i is the total number of infected people (that is daily cases) at day i (with the specimen of their positive test collected at day i), R_i the shedding rate at day i and $g_{i,j}$ the probability of being at the j-th day of the disease at the calendar day i. The governing equations take the form:

$$F_i = \sum_{j=i-m}^{i+n} f_j \qquad (11)$$

$$g_{i,j} = \frac{f_{j-m+i}}{F_i} \qquad (12)$$

$$R_i = \sum_{j=i-m}^{i+n} g_{i,j} S_j \qquad (13)$$

for i>m and i<N-n

where S_j is the shedding rate per person at the j-th day of the disease ($m=\tau_e-\tau_d$, $n=\tau_d$)

RESULTS

Figure 2 displays medical surveillance data reported for the city of Thessaloniki (~ 700,000 inhabitants) (National Public Health Organization, 2020). More specifically, it presents the daily number of infected people versus the date of their specimen collection for medical testing. The date of specimen collection is back-dated by 1 to 4 days (median of 3 days) from the date of reporting by the Hellenic National Public Health Organization. The difference between the two characteristic dates is small but the epidemiological data adjusted to the date of specimen collection are more appropriate to compare with wastewater measurements because they are better associated with the date of infection and the date of symptoms onset. Nevertheless, if early warning by wastewater measurements is the stake then comparisons should be made with medical surveillance data as announced, i.e., based on the date of reporting.

Figure 2. Medical surveillance data reported for the city of Thessaloniki from September 1st, 2020 to January 6th, 2021 and corresponding Bayesian model fit curve.

The presented data cover the period from September, 1st 2020 to January 6th, 2021. A fitting curve is fitted over the raw data to smooth out the noise. Most of the parametric study that follows is performed with respect to this curve, to skip the noise. Following the loose atmosphere of Summer 2020, in September and October there was no strict quarantine in the city and only a modest rule for social distancing was active along with a rule for limited number of people in confined places, like stores and restaurants. In September and the first 10 days of October the daily reported cases were always well below 20 and often even below 10. At around mid-October 2020 the number of infected cases started to escalate exponentially. In just two weeks, the number of cases increased about ten-fold in just two weeks. This dramatic rise was consistent with the about 2.7 days doubling time of the epidemic reported for European countries in the absence of control measures (SET-C Steering Committee, 2020).

373 A strict lock-down was imposed on the city on November 2nd meant to suppress the outburst of the
374 disease. Between Nov 3rd and Nov 17th the number of infected cases fluctuated approximately from 800
375 to 300 cases between weekdays and weekends. Similar fluctuations were noticed also in other studies
376 (Peccia et al., 2020). On December 12th some of the strict measures were released, e.g., retail stores
377 opened to serve people but only at their entrance door and only through pre-scheduled appointments.
378 As a result, after mid-November the number of infected cases gradually went down until it became
379 pretty stable at below 100 daily cases in the first week of 2021.

380 **Figure 3** presents the experimentally determined relative shedding rate of viral RNA copies,
381 $r_1(t)=R_{exp}/R_{expo}$, in sewage from October 5th, 2020 to January 6th, 2021 for Thessaloniki. October 5th is the
382 last day after summer that the measured viral RNA copies in wastewater fluctuated around the
383 experimental limit of quantification of the employed technique (~10 viral RNA copies/mL). R_{exp}
384 represents the daily value of shedding rate, whereas R_{expo} is a reference shedding rate. Presenting
385 experimental shedding data in the form of a ratio reduces the effect of determination uncertainty. R_{expo}
386 has been defined as the average R_{exp} value across the days of the first week of October which was a
387 period with calm epidemiological conditions in the city (less than 10 daily reported cases).

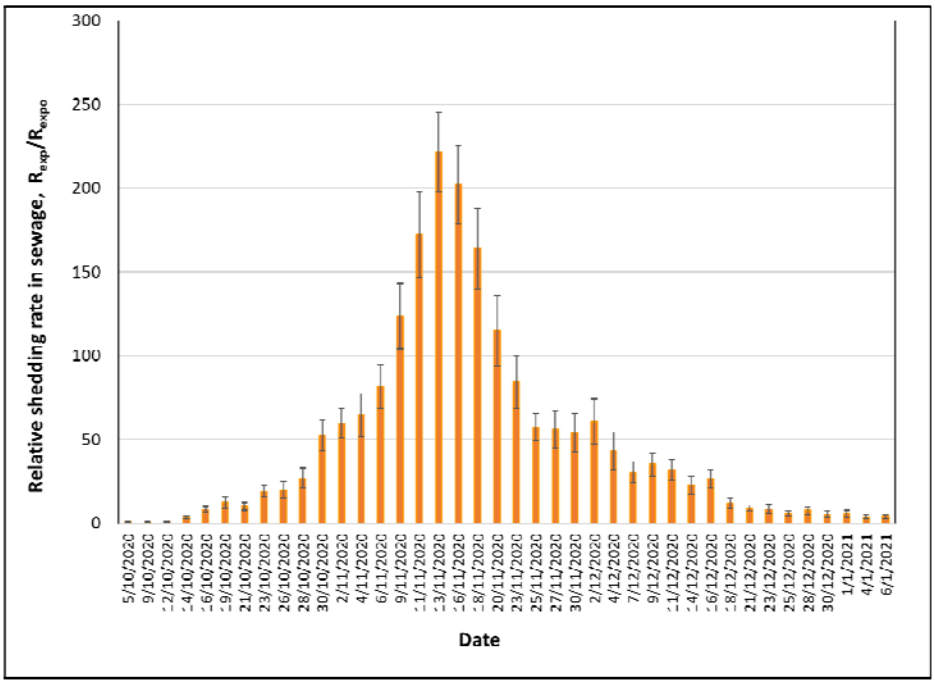


Figure 3. Experimental relative shedding rate in sewage expressed as the ratio of the measured daily shedding rate (R_{exp}) over the reference shedding rate (R_{expo}) from October 5th, 2020 to January 6th, 2021.

There are certain similarities and differences between the medical surveillance data in **Figure 2** and the wastewater data in **Figure 3**. They both show an initial abrupt ascend followed by a later gradual decline. They both show a peak in the first half of November 2020. Yet, the peak in wastewater data is sharp at around November 13th whereas the peak in medical surveillance, despite the intense noise, looks like a plateau between November 3rd and 12th. The initial increasing slope in October's data in these Figures is a bit steeper for wastewater data implying that wastewater surveillance may act as an early warning indicator for medical surveillance. More about this below.

Next, the time series of the reported daily infected cases in the city of Thessaloniki is analyzed using the procedure in the previous section to determine the functions F, g, R. It is reminded that these three functions correspond respectively to (i) the total number of infected people with positive test at a calendar day, t, (those already registered as positive to day t but also those that will be tested in the

402 next days and their test result will be also registered to day t), (ii) the distribution of the number of
403 infected people with respect to the days of the disease, τ, at a calendar day, t, and (iii) the total shedding
404 rate with respect to the calendar day, t. It is important to note that since the temporal discretization
405 quantity is one day, the continuous and the discrete forms of the above functions are arithmetically
406 similar. Thus, the presented results can be read both as continuous functions (per day) or as discrete
407 values (i.e., histograms). The essence of the proposed approach is that it accounts for the time
408 dependence of the shedding rate during the disease. Yet, this is important only when the distribution of
409 the daily shedding rates is not uniform among the population, but it evolves along the days of the
410 disease.

411 Before a parametric analysis is performed it is useful to identify realistic range of values of the model
412 parameters in order to examine their influence. It is reminded that the three model parameters are (i)
413 the total number of shedding days counted from the infection day, τ_e, (ii) the day of the maximum
414 shedding rate, τ_a, and (iii) the day of detection (specimen collection), τ_d. The analysis of the data of
415 Wölfel et al.(Wölfel et al., 2020) by Miura et al(Miura et al., 2021) indicated an average shedding period
416 of 26 days after the onset of symptoms. This is a bit longer than the 17 and 18 days, reported by Huang
417 et al.(Huang et al., 2020) and Tan et al.(Tan et al., 2020), respectively. But both the latter studies
418 reported a high variance in their average shedding period plus they mentioned that in some individuals
419 shedding exceeded five weeks. Apart from this, Tan et al.(Tan et al., 2020), indicated a 6 day median
420 incubation period from the day of infection until the day of symptoms onset. If one considers that the
421 day of symptoms onset is a good candidate for the average day of maximum shedding rate then τ_a= 6.
422 For Greek patients it is realistic to assume that specimen collection for medical testing takes place on
423 the average about 3 days after the onset of symptoms. This is in line with information in literature that
424 the median time period from symptom onset to hospital admission is 3 days(Huang et al., 2020).
425 Therefore, a realistic average day of detection is τ_d=8. The above information combined implies a total

number of shedding days τ_e=32 (6+26) and this is the value adopted herein. Summarizing, the base case parameter values around which parametric analysis is performed are τ_a= 6, τ_d=8 and τ_e=32.

Figure 4 shows the computed values of g(τ,t) for selected values of calendar days, namely, September 25th, October 10th, October 23th, December 9th and December 29th, 2020. The small oscillation appeared is due to the oscillations in the curve f(t). However, as f(t) – the positive tests count density- increases new patients are added to the number of infected people. The addition occurs at a higher rate than the rate of withdrawal of cured people. This leads to the accumulation of infected people at early shedding days, τ, so the function g(τ) acquires the characteristic, decreasing with τ, profiles shown for October 10th and October 23th. In the time regime of the maximum f(t), e.g. curve for December 9th, the function g(τ) tends fast towards uniformity, like at the early shedding days. Finally, during the decreasing period of f(t), e.g. curve for December 29th, the withdrawal of cured people rate overwhelms the entry of new patients so g(t) is an almost linearly increasing function of τ. The above description manifests that the function g(τ) varies a lot with calendar time t and it is far from uniform along the shedding days so the present analysis is necessary for the estimation of shedding rate across calendar days. It is clear that at a particular calendar day, infected people are at a different stage of the disease, and thus, at a different stage of viral shedding.

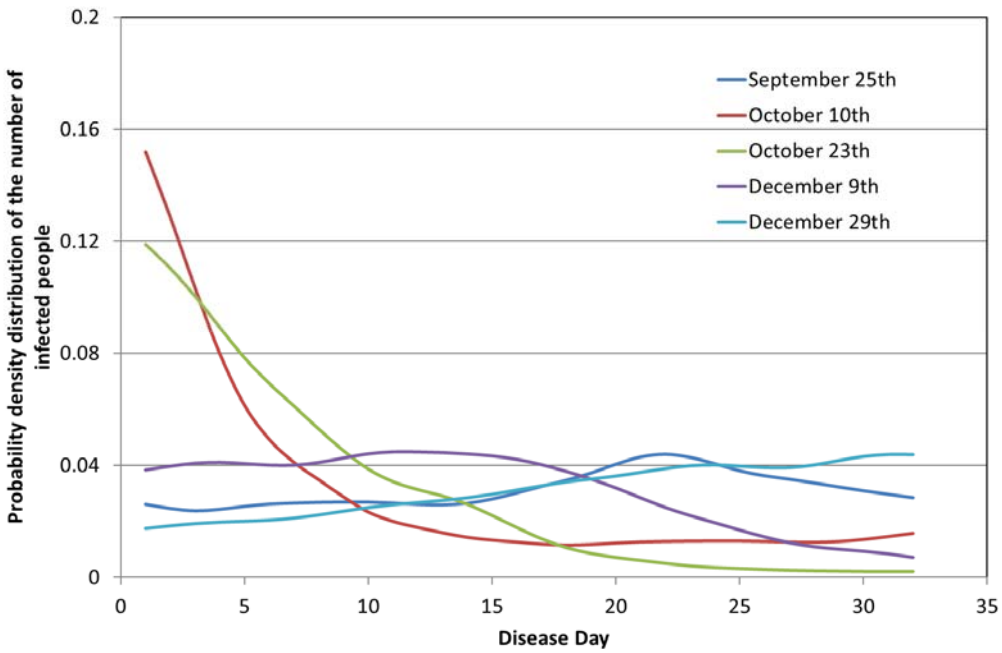
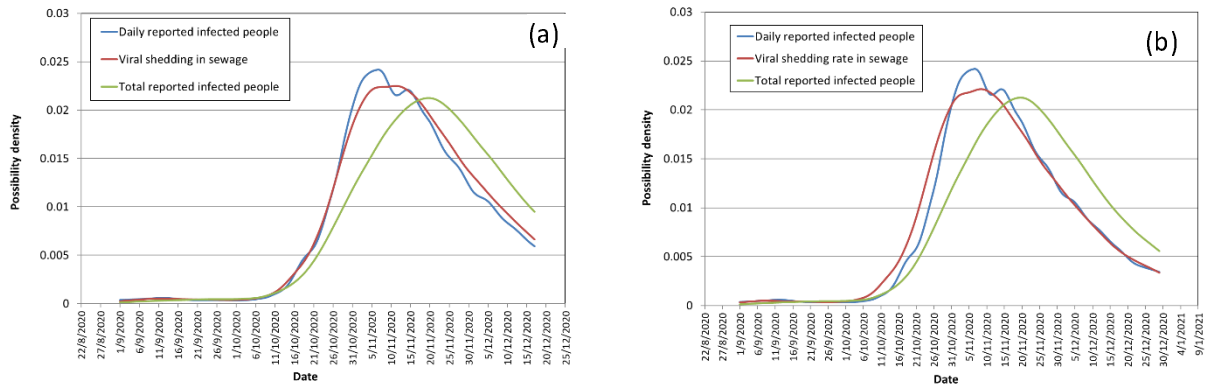


Figure 4. The probability density distribution of the number of infected people with respect to the days of the disease.

The early warning capacity of wastewater surveillance with respect to medical surveillance is estimated by comparing in **Figure 5** the dynamics of functions f(t): daily reported infected people, F(t): total reported infected people, and R(t): experimentally determined viral shedding rate in sewage. More specifically, **Figure 5a** shows this comparison for the base case τ_a= 6, τ_d=8, τ_e=32; **Figure 5b** for τ_a= 2, τ_d=8, τ_e=32; **Figure 5c** for τ_a= 12, τ_d=8, τ_e=32 (all values in days). Curves in **Figure 5** stop at December 29th, 2020, and not at January 6th, 2021. This is because calculations of F(t) -and accordingly of R(t)- can be performed only up to December 29th, 2020, as this function requires for every daily value available data for the subsequent τ_d=8 days. To avoid experimental noise in the parametric analysis, calculations in **Figure 5** are based on the Bayesian fitted curve in **Figure 2**. For the comparison it is imperative to express these three functions in a similar scale. This is done by transforming them in their probability density function counterparts. For the present case, this is done simply by dividing each function with the sum of its values for the period of interest. The function F(t) always resides after f(t) which is typical

for a cumulative type of function. The temporal distance apart between these two functions depends on the values of the parameters τ_d and τ_e. This distance increases as τ_d decreases and τ_e increases. The dynamics of the function R(t) depend also on τ_a. In case of $\tau_a<\tau_d$ then R(t) precedes f(t). This condition is extremely important as it demonstrates that for wastewater surveillance to precede medical surveillance the day of the peak of viral shedding rate in stool during the disease days must lie before the day of specimen collection for medical testing. In all other cases, the curve R(t) resides between f(t) and F(t), with the latter being the approximate long time bound for the dynamics of R(t). The previous statement can be confirmed by observing the three aforementioned functions for the three set of parameters in **Figure 5**. Depending on the calendar day, function R(t) precedes f(t) roughly from 0 to 4 days (**Figure 5b**). For the base case in **Figure 5a** the difference between R(t) and f(t) appears marginal. Yet, one must recall that in **Figure 2** the ascending part of the Bayesian curve precedes raw data, so in reality even for the base case sewage data lie earlier than medical data. In literature there is experimental evidence of identifying SARS-CoV-2 in wastewater earlier than medical reporting by several days, e.g., 2 days in Ottawa Canada(D'Aoust et al., 2021), 2-4 days in Montana USA(Nemudryi et al., 2020). In summary, the viral shedding rate evolution curve measured in wastewater lies from a few days before the curve of the number of daily infected people up to the curve of the total number of infected people. The exact position of the curves depends on the relation between the day of maximum shedding rate in stool and the day of specimen collection for medical testing.



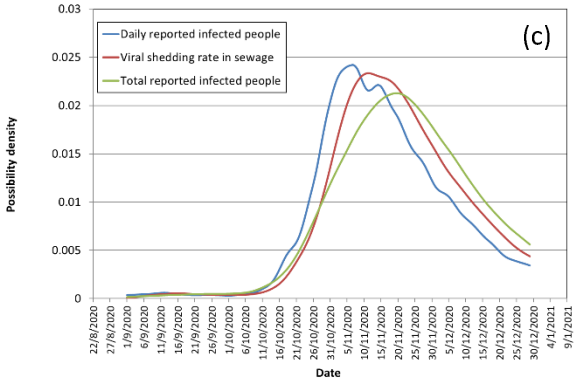
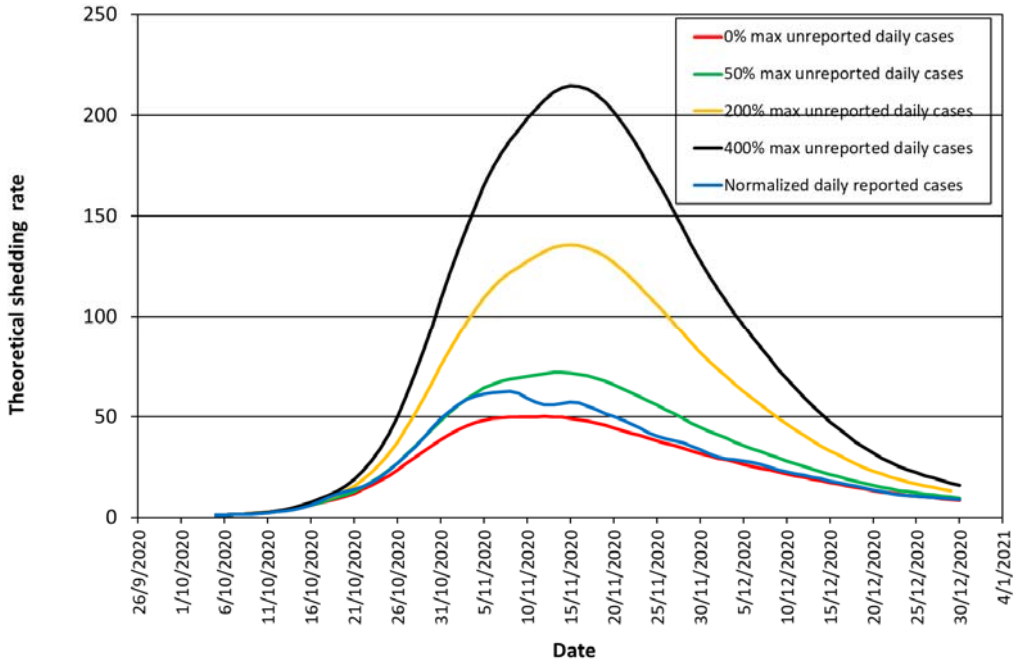


Figure 5. Comparison of the relative viral shedding rate in sewage with the number of the daily and total reported infected people per surveillance day. In all cases shedding duration in stool is set as 32days, laboratory test is assumed at DAY 8, whereas the day of shedding peak in stool is at: DAY 6 (a), DAY 2 (b) and DAY 12 (c).

For appraising in real life the early diagnostic capacity of wastewater surveillance it is important to compare wastewater data with medical surveillance data based on the date of reporting (and not on the date of specimen collection). Peccia et al.(Peccia et al., 2020) found in New Haven (USA) that although wastewater data was ahead by only 0-2 days of positive test results by the date of specimen collection, it was 6-8 days ahead of positive test results by the reporting date. In Thessaloniki (National Public Health Organization, 2020) the date of reporting follows the date of specimen collection by about 2 days on the average and so even the base case parameters in **Figure 5a** permit an early diagnosis capacity of wastewater surveillance of around 2 days.

A critical issue in the disease dynamics is the effect of the number of unreported infected people (unreported cases). Let us denote as U the ratio of unreported to reported infected people. If U is constant along the days of the shedding period, the analysis is exactly the same with that shown above for the reported infected people, F, since the probability density functions do not change. However, if the number of unreported people varies along the shedding period, then this may yield a time lag between wastewater and medical data. **Figure 6** compares the relative medical surveillance data of the

493 smoothed (Bayesian fit) daily reported cases, $f(t)/f_o$, versus the theoretically estimated relative shedding
494 rate in wastewater, $r_2(t)=R(t)/R_o$. Normalization parameters, f_o and R_o are average values of the
495 respective parameters over the same reference first week of October. Comparisons are for the base
496 case τ_a= 6, τ_d=8, τ_e=32. The ratio U is assumed to vary proportionally with the number of reported
497 infected people. The parameter U_{max} is used to parameterize U designating the maximum value of U for
498 the case considered. The parameter U_{max} takes the value 0.5, 2 and 4 in **Figure 6**. Apparently, for
499 U_{max}=0.5 (unreported cases becomes at most 50% of reported cases) wastewater and medical data
500 coincide at the ascent of the curves but at the descent wastewater lags behind. This changes for U_{max}=2
501 (unreported cases becomes at most 200% of reported cases) and for U_{max}=4 (unreported cases becomes
502 at most 400% of reported cases) where wastewater data clearly go before medical data at the ascent of
503 the curve. Summing up, a rising number of unreported cases as the shedding rate increases leads to
504 earlier wastewater signal than medical surveillance. Furthermore, the higher the maximum number of
505 unreported cases the more in advance the wastewater data from the medical data.



506

507 **Figure 6.** Comparison of the theoretical relative shedding rate in sewage with the relative medical
508 surveillance data of the smoothed (Bayesian fit) daily reported cases for different scenarios of
509 unreported daily cases. Unreported daily cases vary proportionally with the number of reported daily
510 cases starting from 0 and reaching a maximum value of 0, 50, 200, 400% of reported cases at the date of
511 maximum shedding rate.

512

513 Next, the experimental relative shedding rate, $r_1(t)=R_{exp}(t)/R_{expo}$, shown in **Figure 3** is compared with the
514 estimated total number of infected people, including both reported and unreported ones, **Figure 7.** Red
515 circles denote increasing shedding rates whereas green squares decreasing shedding rates, respectively.
516 It is apparent that from low to moderate relative shedding rates ($R_{exp}(t)/R_{expo}$ <100) and total number of
517 infected people (F<15000) there is a roughly linear relationship between the two quantities. This
518 changes dramatically for higher values of either quantities. Moreover, during the increasing phase of the
519 relative shedding rate, i.e., outbreak of pandemic wave, the same number of infected people sheds a
520 higher viral load than during the decreasing phase. This is so because at the outbreak phase of the
521 pandemic most infected people are at the early phase of their disease when shedding rate is higher, as
522 shown in **Figure 1**. In line with this, when the relative shedding rate starts to fall right after its peak, the
523 total number of infected people continues to rise to about 10% more cases. This happens because the
524 turning point (peak) of the curve designates the moment when the daily new infected cases begin to
525 reduce, **Figure 2**, so the majority of the total number of infected people are at later phase of their
526 disease when shedding rate is lower.

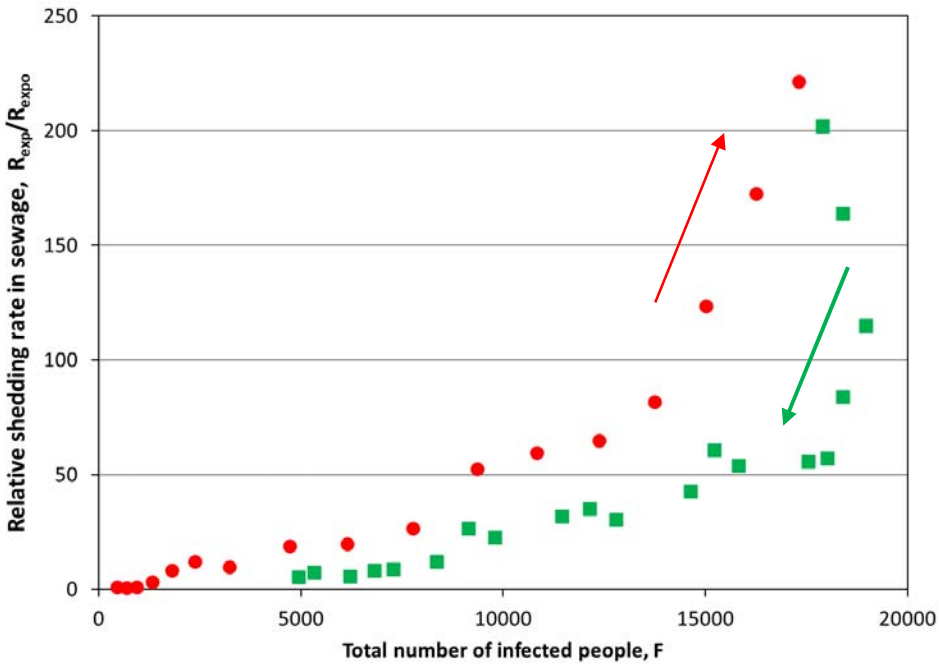


Figure 7. Experimental relative shedding rate in sewage, $R_{exp}(t)/R_{expo}$, versus the estimated total number of infected people (reported and unreported), F. The latter is estimated assuming that, on the average, the shedding duration in stool is 32 days and medical tests are taken at DAY 8 during the disease. Red circles denote increasing shedding rates whereas green squares denote decreasing shedding rates.

Let us now try to correlate the theoretically derived function R(t) with the actual one found from experimental wastewater analysis $R_{exp}(t)$. The latter was shown as a relative shedding rate, $r_1(t)=R_{exp}(t)/R_{expo}$, in **Figs 3** and **7**. Contrary to what was done in **Figure 6**, now the theoretically estimated relative shedding rate $r_2(t)=R(t)/R_o$ is computed from the noisy medical surveillance raw data in **Figure 2** for the base case parameter values τ_a= 6, τ_d= 8 and τ_e=32. As a result, $r_2(t)$ data are also noisy. Therefore, $\beta(t)=r_1(t)/r_2(t)$ is a ratio reflecting the total number of reported and unreported infected people, F(1+U), over the total number of reported infected people, F, at every calendar day. In other words, β=1+U, that is, β-1 stands for U, the ratio of the total unreported over the total reported cases. Thus, the minimum value of β is 1 and corresponds to zero unreported cases. As before, the values of R_{expo} and R_o both refer to the same reference first week of October. Selection of the reference week at an

epidemiologically calm period allows assuming that $R_{expo}/R_o \approx 1$ so the normalization in β can be ignored. The evolution of β with the calendar days appears in **Figure 8.** Considering the intense day-by-day scatter, the ratio β takes values roughly between 1 and 5, following a qualitatively similar trend with the wastewater shedding data in **Figure 3**. Interestingly, these results imply that there are essentially no unreported cases at periods of low shedding but unreported cases rise up to four times the reported ones (U/F≈4) when shedding is maximum. Analysis of seroprevalence data in Qatar indicated that diagnosed infections represented about 10% of actual cases(Saththasivam et al., 2021). Another serological-testing study(Havers et al., 2020) showed that the actual number of infections could be from 6 to 24 times the number of the reported cases. Overall, the presently estimated values of U/F from ~0 to 4 is close to the above published values. To our knowledge, this is the first time in the SARS-CoV-2 pandemic that wastewater data are employed to indicate a proportion between unreported and reported infected cases. Moreover, the present study shows that the proportion of unreported to reported cases is not constant during the outbreak of a disease wave but as the number of infected people (and shedding rate) increases, it attains higher values and its scatter decreases.

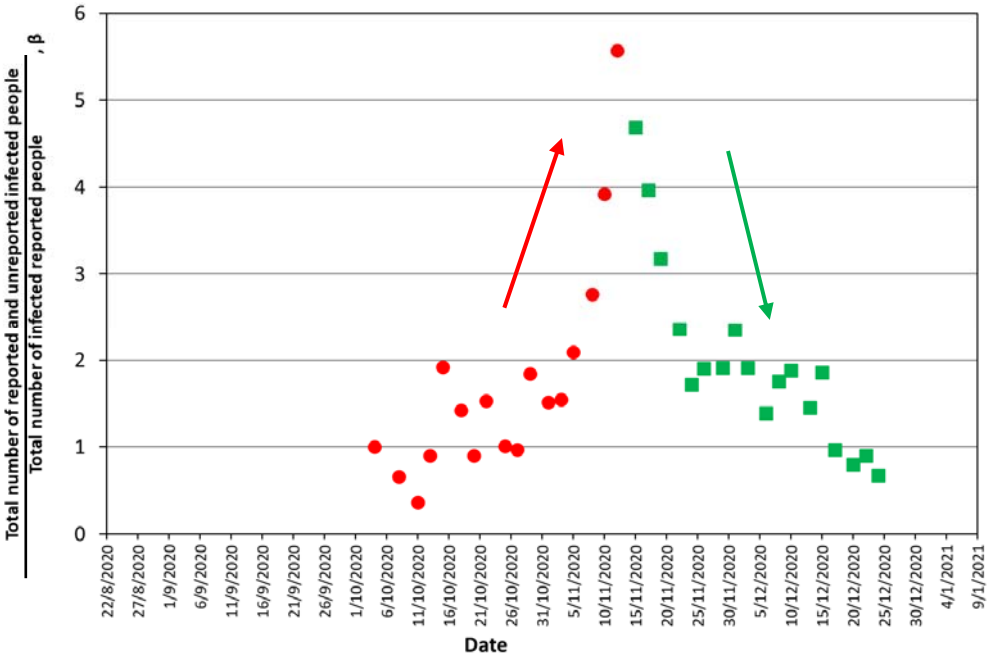


Figure 8. Evolution of the ratio of the total number of infected people (reported + unreported) over the number of reported infected people. The ratio is calculated from unsmoothed raw data in Figures 2 and 3. Red circles denote increasing shedding rates whereas green squares denote decreasing shedding rates.

Based on the above, an effort is made to correlate β to the total number of the reported infected cases F. The result is shown in **Figure 9a** where once more red circles denote increasing shedding rates and green squares denote decreasing shedding rates. Alike in **Figure 7**, β starts decreasing before the maximum F is reached. In addition, at high F values there is a hysteresis in β between increasing and decreasing branches of the curves. This might imply that virus spread among unreported people is a bit higher in the declining phase of the disease either because of the cumulatively higher virus prevalence in the community or because more people undergo testing during the ascending spreading period of the disease than in the milder descending period. **Figure 9b** presents β versus the daily reported number of the infected cases, f. Interestingly there is an almost linear relation between the two parameters up to about 400 daily reported cases. After that point, spreading of the virus in the community is too high so β depends on the total number infected people F and not on the daily reported cases.

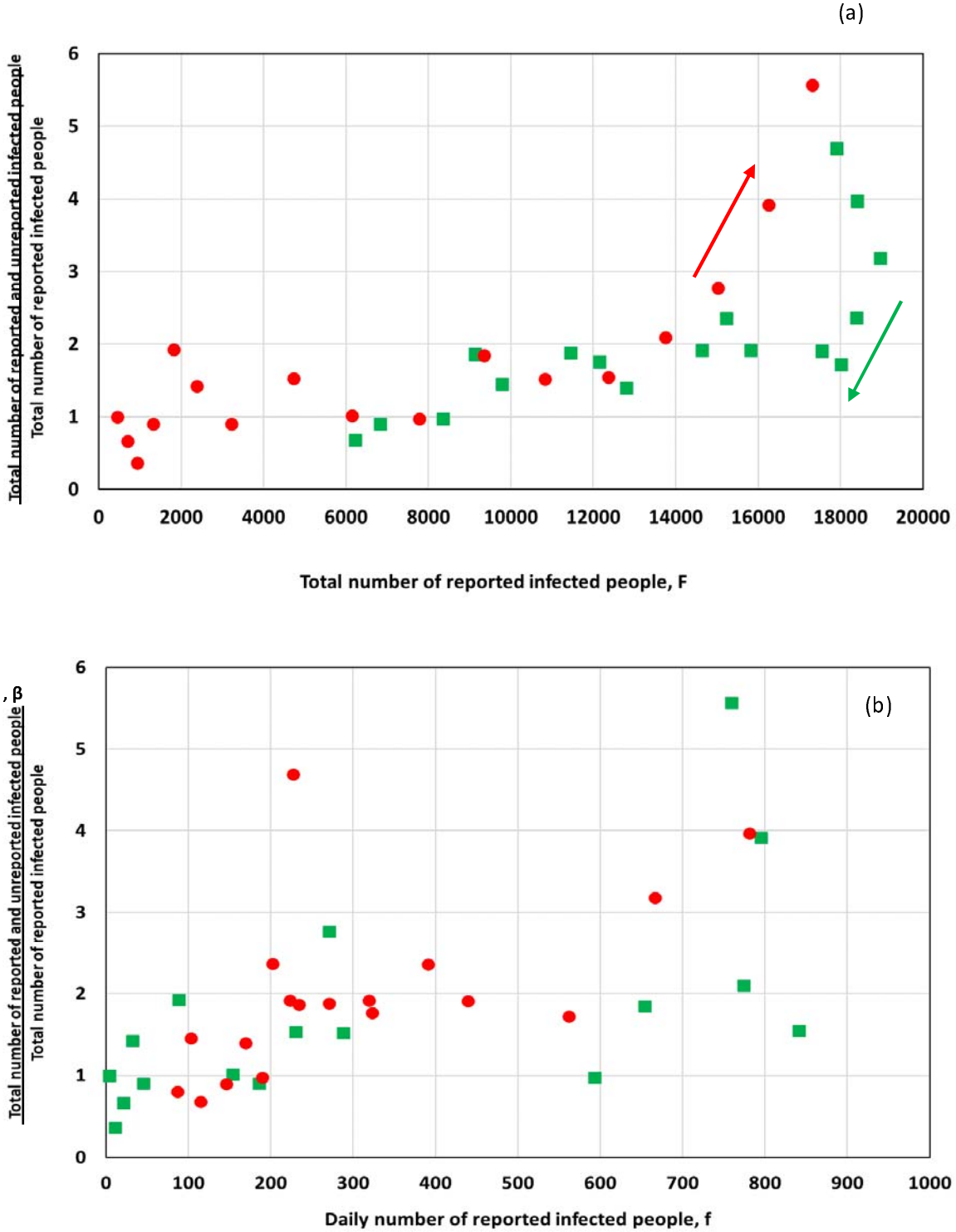


Figure 9. Variation of the ratio of the total number of infected people (reported + unreported) over the number of reported infected people with regards to (a) the total number of reported infected people

575 and (b) the daily number of infected people. The ratio is calculated from unsmoothed raw data in
576 Figures 2 and 3. Red circles denote increasing shedding rates whereas green squares denote decreasing
577 shedding rates.

578 It must be mentioned that β may depend also on other quantities (apart from F and the
579 ascending/descending mode), like the value of the F slope, but at present there are not enough data to
580 support this. Having an estimation of the total number of infected people through β, a crude estimation
581 of the product AB (average maximum virus shedding rate per person) is obtained. The estimated value
582 of AB is at about $2 \cdot 10^{11}$ day^{-1}. This value is in fair proximity with that estimated by the wastewater
583 analysis of Wu et al.(Wu et al., 2020). However, more work is needed to allow judging on the
584 correctness of this value.

585

586 CONCLUSIONS

587 A mathematical model is developed that permits estimation of SARS-CoV-2 shedding rate in wastewater
588 from the number of daily infected people (daily cases) announced by medical surveillance. The problem
589 is complex as it requires a cumulative function of the total number of infected people at every calendar
590 day and a function of the average shedding rate among infected individuals at every day along the
591 course of the disease. Based on the limited available evidence in literature, a realistic function for the
592 shedding rate of SARS-CoV-2 in stool of infected individuals is proposed which calls for an exponential
593 increase of shedding rate from the day of infection to the day of symptoms onset, being followed by an
594 exponential decay until the end of the disease. Three characteristic times describe this function: the day
595 of maximum shedding rate, the day of detection (specimen collection) and the end day after the onset
596 of the disease. Applying this model to the public health surveillance data for the city of Thessaloniki
597 (~700,000 inhabitants, North Greece) a thorough parametric study is performed. It is shown that for

WBE to afford an early warning capacity, the day of maximum shedding rate must precede the day of detection by a few days. In particular, at the beginning of an outbreak curve, a 4-day early WBE signal requires these two characteristic times to be apart by 6 days. For Thessaloniki where on the average these two characteristic times are apart by just 2 days but, additionally, the day of reporting follows the date of specimen collection also by about 2 days, the early diagnosis capacity of wastewater surveillance is around 2 days. Furthermore, comparison of wastewater surveillance and public health data indicates the existence of a number of unreported infected people. For Thessaloniki in the outbreak of November 2020, this number was negligible at epidemiologically calm days with low shedding rate but went up to four times the number of reported people when shedding rate reached a climax at mid-November 2020. Interestingly, the presence of an increasing number of unreported cases with shedding rate enhances the early warning capacity of WBE. To this end, the present model is an essential tool to investigate the dynamics of virus spreading based on wastewater measurements. When such knowledge is adequately acquired then the inverse problem of estimating the number of cases from wastewater data can be attempted.

REFERENCES

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tscharke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. v., Mueller, J.F., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. Science of the Total Environment 728, 138764. https://doi.org/10.1016/j.scitotenv.2020.138764

619 CDC, 2021. CDC [WWW Document]. Center for Disease Control and Prevention, CDC. URL
620 https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/wastewater-surveillance/public-
621 health-interpretation.html (accessed 3.10.21).

622 Chaintoutis, S.C., Papadopoulou, E., Melidou, A., Papa, A., Dovas, C.I., 2019. A PCR-based NGS protocol
623 for whole genome sequencing of West Nile virus lineage 2 directly from biological specimens.
624 Molecular and Cellular Probes 46, 101412. https://doi.org/10.1016/j.mcp.2019.06.002

625 Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K.W., Bleicker, T., Brünink, S.,
626 Schneider, J., Schmidt, M.L., Mulders, D.G.J.C., Haagmans, B.L., van der Veer, B., van den Brink, S.,
627 Wijsman, L., Goderski, G., Romette, J.L., Ellis, J., Zambon, M., Peiris, M., Goossens, H., Reusken, C.,
628 Koopmans, M.P.G., Drosten, C., 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-
629 time RT-PCR. Eurosurveillance 25. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

630 D'Aoust, P.M., Mercier, E., Montpetit, D., Jia, J.J., Alexandrov, I., Neault, N., Baig, A.T., Mayne, J., Zhang,
631 X., Alain, T., Langlois, M.A., Servos, M.R., MacKenzie, M., Figeys, D., MacKenzie, A.E., Graber, T.E.,
632 Delatolla, R., 2021. Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in
633 communities with low COVID-19 incidence and prevalence. Water Research 188, 116560.
634 https://doi.org/10.1016/j.watres.2020.116560

635 Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M.H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J.,
636 Gonzalez, D., 2020. COVID-19 surveillance in Southeastern Virginia using wastewater-based
637 epidemiology. Water Research 186, 116296. https://doi.org/10.1016/j.watres.2020.116296

638 Guan, W., Ni, Z., Hu, Yu, Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D.S.C., Du, B., Li, L., Zeng,
639 G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, Jin-lin, Liang, Z., Peng,
640 Y., Wei, L., Liu, Y., Hu, Ya-hua, Peng, P., Wang, Jian-ming, Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S.,

641 Luo, J., Ye, C., Zhu, S., Zhong, N., 2020. Clinical Characteristics of Coronavirus Disease 2019 in China.
642 New England Journal of Medicine 382, 1708–1720. https://doi.org/10.1056/nejmoa2002032

643 Havers, F.P., Reed, C., Lim, T., Montgomery, J.M., Klena, J.D., Hall, A.J., Fry, A.M., Cannon, D.L., Chiang,
644 C.F., Gibbons, A., Krapiunaya, I., Morales-Betoulle, M., Roguski, K., Rasheed, M.A.U., Freeman, B.,
645 Lester, S., Mills, L., Carroll, D.S., Owen, S.M., Johnson, J.A., Semenova, V., Blackmore, C., Blog, D.,
646 Chai, S.J., Dunn, A., Hand, J., Jain, S., Lindquist, S., Lynfield, R., Pritchard, S., Sokol, T., Sosa, L.,
647 Turabelidze, G., Watkins, S.M., Wiesman, J., Williams, R.W., Yendell, S., Schiffer, J., Thornburg, N.J.,
648 2020. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May
649 12, 2020. JAMA Internal Medicine 180, 1776–1786.
650 https://doi.org/10.1001/jamainternmed.2020.4130

651 He, X., Lau, E.H.Y., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y.C., Wong, J.Y., Guan, Y., Tan, X., Mo, X.,
652 Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., Zhang, F., Cowling, B.J.,
653 Li, F., Leung, G.M., 2020. Temporal dynamics in viral shedding and transmissibility of COVID-19.
654 Nature Medicine 26, 672–675. https://doi.org/10.1038/s41591-020-0869-5

655 Huang, J., Mao, T., Li, S., Wu, L., Xu, X., Li, H., Xu, C., Su, F., Dai, J., Shi, J., Cai, J., Huang, C., Lin, X., Chen,
656 D., Lin, X.L., Sun, B., Tang, S., 2020. Long period dynamics of viral load and antibodies for SARS-CoV-
657 2 infection: An observational cohort study. medRxiv.
658 https://doi.org/10.1101/2020.04.22.20071258

659 la Rosa, G., Iaconelli, M., Mancini, P., Bonanno Ferraro, G., Veneri, C., Bonadonna, L., Lucentini, L.,
660 Suffredini, E., 2020. First detection of SARS-CoV-2 in untreated wastewaters in Italy. Science of the
661 Total Environment 736. https://doi.org/10.1016/j.scitotenv.2020.139652

662 Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J.,
663 2020. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported
664 confirmed cases: Estimation and application. Annals of Internal Medicine 172, 577–582.
665 https://doi.org/10.7326/M20-0504

666 Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing,
667 X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R.,
668 Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu,
669 B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M.,
670 Feng, Z., 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected
671 Pneumonia. New England Journal of Medicine 382, 1199–1207.
672 https://doi.org/10.1056/nejmoa2001316

673 Medema, G., Been, F., Heijnen, L., Petterson, S., 2020a. Implementation of environmental surveillance
674 for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. Current
675 Opinion in Environmental Science and Health 17, 49–71.
676 https://doi.org/10.1016/j.coesh.2020.09.006

677 Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020b. Presence of SARS-Coronavirus-2
678 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the
679 Epidemic in The Netherlands. Environmental Science & Technology Letters 7, 511–516.
680 https://doi.org/10.1021/acs.estlett.0c00357

681 Miura, F., Kitajima, M., Omori, R., 2021. Duration of SARS-CoV-2 viral shedding in faeces as a parameter
682 for wastewater-based epidemiology: Re-analysis of patient data using a shedding dynamics model.
683 Science of the Total Environment 769, 144549. https://doi.org/10.1016/j.scitotenv.2020.144549

684 National Public Health Organization, 2020. Daily reports COVID-19 2020. URL
685 https://eody.gov.gr/epidimiologika-statistika-dedomena/ektheseis-covid-19/imerisies-ektheseis-
686 covid-19-2020/ (accessed 5.14.21).

687 Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K.K.,
688 Wilkinson, R., Wiedenheft, B., 2020. Temporal Detection and Phylogenetic Assessment of SARS-
689 CoV-2 in Municipal Wastewater. Cell Reports Medicine 1, 100098.
690 https://doi.org/10.1016/j.xcrm.2020.100098

691 Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik,
692 A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D.M., Arnold, W., Omer, S.B., 2020.
693 Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nature
694 Biotechnology 38, 1164–1167. https://doi.org/10.1038/s41587-020-0684-z

695 Petala, M., Dafou, D., Kostoglou, M., Karapantsios, T., Kanata, E., Chatziefstathiou, A., Sakaveli, F.,
696 Kotoulas, K., Arsenakis, M., Roilides, E., Sklaviadis, T., Metallidis, S., Papa, A., Stylianidis, E.,
697 Papadopoulos, A., Papaioannou, N., 2021. A physicochemical model for rationalizing SARS-CoV-2
698 concentration in sewage. Case study: The city of Thessaloniki in Greece. Science of the Total
699 Environment 755, 142855. https://doi.org/10.1016/j.scitotenv.2020.142855

700 Saththasivam, J., El-Malah, S.S., Gomez, T.A., Jabbar, K.A., Remanan, R., Krishnankutty, A.K., Ogunbiyi,
701 O., Rasool, K., Ashhab, S., Rashkeev, S., Bensaad, M., Ahmed, A.A., Mohamoud, Y.A., Malek, J.A.,
702 Abu Raddad, L.J., Jeremijenko, A., Abu Halaweh, H.A., Lawler, J., Mahmoud, K.A., 2021. COVID-19
703 (SARS-CoV-2) outbreak monitoring using wastewater-based epidemiology in Qatar. Science of the
704 Total Environment 774, 145608. https://doi.org/10.1016/j.scitotenv.2021.145608

705 Sellaoui, L., Badawi, M., Monari, A., Tatarchuk, T., Jemli, S., 2020. Since January 2020 Elsevier has
706 created a COVID-19 resource centre with free information in English and Mandarin on the novel
707 coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company
708 ' s public news and information .

709 SET-C Steering Committee, 2020. Reproduction number (R) and growth rate (r) of the COVID-19
710 epidemic in the UK 1–86.

711 Siegel, A., Chang, P.J., Jarou, Z.J., Paushter, D.M., Harmath, C.B., Arevalo, J. ben, Dachman, A., 2020.
712 Lung base findings of coronavirus disease (COVID-19) on abdominal CT in patients with
713 predominant gastrointestinal symptoms. American Journal of Roentgenology 215, 607–609.
714 https://doi.org/10.2214/AJR.20.23232

715 Tan, W., Lu, Y., Zhang, J., Wang, J., Dan, Y., Tan, Z., He, X., Qian, C., Sun, Q., Hu, Q., Liu, H., Ye, S., Xiang,
716 X., Zhou, Y., Zhang, W., Guo, Y., Wang, X.H., He, W., Wan, X., Sun, F., Wei, Q., Chen, C., Pan, G., Xia,
717 J., Mao, Q., Chen, Y., Deng, G., 2020. Viral kinetics and antibody responses in patients with COVID-
718 19. medRxiv. https://doi.org/10.1101/2020.03.24.20042382

719 Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones,
720 T.C., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R.,
721 Zwirglmaier, K., Drosten, C., Wendtner, C., 2020. Virological assessment of hospitalized patients
722 with COVID-2019. Nature 581, 465. https://doi.org/10.1038/s41586-020-2196-x

723 Wu, F., Zhang, J., Xiao, A., Gu, X., Lee, W.L., Armas, F., Kauffman, K., Hanage, W., Matus, M., Ghaeli, N.,
724 Endo, N., Duvallet, C., Poyet, M., Moniz, K., Washburne, A.D., Erickson, T.B., Chai, P.R., Thompson,
725 J., Alm, E.J., 2020. SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically
726 Confirmed Cases. mSystems 5. https://doi.org/10.1128/mSystems.00614-20

727 Wurtzer[1], S., Marechal, V., Mouchel[4], J.M., Maday, Y., Teyssou[6], R., Richard[1], E., Almayrac[7], J.L., Moulin[1],
728 L., n.d. Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification
729 in waste water 1. https://doi.org/10.2807/1560-7917

730 Ye, Y., 2018. The Detection and Fate of Enveloped Viruses in Water Environments 149.

731 Zheng, S., Fan, J., Yu, F., Feng, B., Lou, B., Zou, Q., Xie, G., Lin, S., Wang, R., Yang, X., Chen, W., Wang, Q.,
732 Zhang, D., Liu, Y., Gong, R., Ma, Z., Lu, S., Xiao, Y., Gu, Y., Zhang, J., Yao, H., Xu, K., Lu, X., Wei, G.,
733 Zhou, J., Fang, Q., Cai, H., Qiu, Y., Sheng, J., Chen, Y., Liang, T., 2020. Viral load dynamics and
734 disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March
735 2020: Retrospective cohort study. The BMJ 369. https://doi.org/10.1136/bmj.m1443

736

737